stochastic processes and their applications

ELSEVIER Stochastic Processes and their Applications 56 (1995) 117–132

On the maximum entropy principle for a class of stochastic processes

Benedikt Horsthemke[a], Markus Rüttermann[b,*]

[a] Allianz GD, München, Germany
[b] Inst. Angew. Mathematik, Universität Bonn, Bonn, Germany

Received September 1992; revised July 1994

Abstract

This paper extends results of Bolthausen and Schmock on the asymptotical behaviour of certain Laplace-type transformations of Markov chains in two aspects: First we consider transformations of a more general class of processes which satisfy an Orey-type fading condition and secondly we study transformations on process level.

AMS 1991 Subject Classifications: 60F10, 60F05, 60K35

Keywords: Maximum entropy principle, Large deviations; Laplace's method; Conditional limit theorem; Empirical process; Conditional expectation; Weak convergence; Fading condition

1. Introduction

Let P be a measure on $(\Omega, \mathscr{F})$, where $\Omega = S^{\mathbb{N}_0}$, S is a Polish space and $\mathscr{F} = \mathscr{B}(\Omega)$ denotes the Borel-σ-field to the product topology on Ω. Let L_n denote the cyclic empirical process.

For a measurable functional $F: \mathscr{M}_1(\Omega) \to [-\infty, \infty)$, which is bounded above, we define the Laplace-type-transformed process $\hat{P}_n$ for $n \in \mathbb{N}$ and $A \in \mathscr{F}$ by

$$\hat{P}_n[A] := \frac{\int 1_A \cdot \exp(n \cdot (F(L_n)) \mathrm{d}P}{\int \exp(n \cdot F(L_n)) \mathrm{d}P}.$$

If the sequence $(P \circ L_n^{-1})_{n \in \mathbb{N}}$ satisfies a large deviation principle with rate function I, one can derive results on the asymptotical behaviour of the sequence $(\hat{P}_n)_{n \in \mathbb{N}}$ under certain regularity assumptions on the measure P and the functional F. We show that under these assumptions the sequence $(\hat{P}_n)_{n \in \mathbb{N}}$ is tight in $\mathscr{M}_1(\Omega)$ and every accumulation point is a mixture of transformed measures of the compact set of those measures

* Corresponding author.

SSDI 0304-4149(94)00060-3

which maximize the functional $(\bar{F} - I)$, where $\bar{F}$ denotes the upper semicontinuous version of F.

This result (Theorem 2.3) has been inspired by the work of Bolthausen and Schmock (1989) and is an extension of their Theorem 2.4 in two directions: the first is that we study transformations on "level-(III)", i.e. we study transformations where the functional F has the full empirical process as its argument and the second and more important extension is that we not only consider transformations of Markov chains, but also transformations of more general measures which satisfy a certain "fading-condition" (Condition 2.1) in the sense that the dependence of the evolution of the process on its states in the past decreases sufficiently fast with their distance in time. A similar, slightly weaker fading-condition is contained in Condition 4.1 in Orey (1984). Orey shows that for measures of this kind (assuming certain additional regularity conditions) a Shannon–McMillan-type theorem is valid and that the sequence $(P \circ L_n^{-1})_{n \in \mathbb{N}}$ satisfies a large deviation principle with the specific relative entropy as rate function. Hence for such measures our Theorem 2.3 applies with the specific relative entropy as rate function.

Our main result also extends Theorem 4.1.16 of Schmock (1990) in the discrete time case, for which Schmock considers transformations of Markov chains where F has any finite dimensional marginals of L_n as its argument. But Schmock also discusses the case of continuous time markov processes, which is not covered by our work.

Georgii (1993) considers transformations of random fields (which we do not), but he obtains a result similar to our Theorem 2.3 only for P being a shift-invariant product measure (Georgii, 1993, Theorem 1.6).

A special case of our theorem is $F = -\infty 1_{\Gamma^c}$, for some $\Gamma \subseteq \mathcal{M}_1(\Omega)$. In this case $\hat{P}_n[A] = P[A \mid L_n \in \Gamma]$, and if P does not belong to the closure of Γ one derives with the help of Theorem 2.3 asymptotic results for the process P under large deviations of the empirical process L_n.

Some papers consider the special case of limit laws for conditional distributions, conditioned on the empirical measure: Csiszar (1984) derives strong convergence results for the process $P(\cdot \mid \rho_n \in \Gamma)$, where P is a stationary product measure, ρ_n denotes the empirical measure and Γ is a convex set of measures. Csiszar et al. (1987) have studied Markov chains on a finite state space, conditioned on the empirical pair measure, using combinatorial methods.

2. Notations and statement of the result

Let S be a Polish space. We equip the product space $\Omega := S^{\mathbb{N}_0}$ with the product topology, so it becomes a Polish space, too. For $n \in \mathbb{N}_0$ and $\omega = (\omega(0), \omega(1), \dots) \in \Omega$ the projection $X_n : \Omega \to S$ is defined by $X_n(\omega) := \omega(n)$.

We denote by $\mathscr{B}(S)$ the Borel-σ-field on S and by $\mathscr{F} := \mathscr{B}(\Omega)$ the Borel-σ-field on Ω. We write $\mathscr{F}_n$ for the σ-field that is generated by the events which occur up to time n, i.e. $\mathscr{F}_n := \sigma(X_0, \dots, X_n)$.

Let $\mathscr{M}_1(\Omega)$, the space of probability measures on $(\Omega, \mathscr{F})$, be equipped with the weak topology which turns it into a Polish space. In the sequel all topological statements concerning probability measures will refer to the weak topology. In particular we mean by convergence of probability measures their weak convergence.

The shift-transformation $T: \Omega \to \Omega$ is defined by $(T\omega)(n) := \omega(n+1)$. We denote by $\mathscr{M}_1^S(\Omega)$ the closed and convex set of stationary probability measures on Ω, i.e. the set of measures that are invariant w.r.t. the shift T:

$$\mathscr{M}_1^S(\Omega) := \{Q \in \mathscr{M}_1(\Omega): Q \circ T^{-1} = Q\}.$$

We define the projection π_k of a measure $Q \in \mathscr{M}_1(\Omega)$ on its marginal distribution on S^k by $\pi_k Q := Q \circ (X_0, \dots, X_{k-1})^{-1}$. The "cyclic" empirical process $L_n: \Omega \to \mathscr{M}_1(\Omega)$ is defined as

$$L_n(\omega) := \frac{1}{n+1} \sum_{k=0}^{n} \delta_{T_n^k(\omega_n)},$$

where

$$T_n(\omega)(i) := \begin{cases} \omega(i+1) & \text{if } i \not\equiv n \bmod n+1, \\ \omega(i-n) & \text{if } i \equiv n \bmod n+1 \end{cases}$$

denotes the cyclic shift and ω_n denotes the periodical continuation of $(\omega(0), \dots, \omega(n)) \in S^{n+1}$ on $\omega_n := (\omega(0), \dots, \omega(n), \omega(0), \dots, \omega(n), \dots) \in \Omega$. Note that $L_n(\omega) \in \mathscr{M}_1^S(\Omega)$ for all $\omega \in \Omega$ and that L_n is $\mathscr{F}_n$-measurable.

Let $P \in \mathscr{M}_1(\Omega)$ be a measure on the path space and let $F: \mathscr{M}_1(\Omega) \to [-\infty, \infty)$ be a measurable functional which is bounded above. If $P[F(L_n) = -\infty] < 1$ for all sufficiently large $n \in \mathbb{N}$ one can define the sequence $(\hat{P}_n)_{n\in\mathbb{N}}$ of Laplace-type transformations of P as

$$\hat{P}_n[A] := \frac{E[1_A \cdot \exp(n \cdot F(L_n))]}{E[\exp(n \cdot F(L_n))]},$$

where $A \in \mathscr{F}$ and E denotes expectation w.r.t. P.

We will now study the asymptotical behaviour of the sequence $(\hat{P}_n)_{n\in\mathbb{N}}$ under suitable assumptions on the process measure P and on the functional F.

Let $\mu \in \mathscr{M}_1(S)$ be the initial distribution of the process and $(K_n)_{n\in\mathbb{N}}$ a family of stochastic kernels, with $K_n: S^n \times \mathscr{B}(S) \to [0, 1]$ for $n \in \mathbb{N}$. Thus, a measure

$$P := \mu \otimes K_1 \otimes K_2 \otimes \cdots \in \mathscr{M}_1(\Omega)$$

is induced, on which we put the following assumption:

Condition 2.1

1. There exists a measure $\pi \in \mathscr{M}_1(S)$, such that for all $n \in \mathbb{N}$

$$K_n(x_0, \dots, x_{n-1}, \mathrm{d}y) = k_n(x_0, \dots, x_{n-1}, y)\pi(\mathrm{d}y) \quad \text{for all } (x_0, \dots, x_{n-1}) \in S^n,$$

and k_n is continuous, bounded and bounded away from zero.

The initial distribution satisfies

$$\mu(\mathrm{d}x) = m(x)\pi(\mathrm{d}x)$$

and m is bounded and continuous.

2. For all $n \in \mathbb{N}$ let α_n be given by

$$1 + \alpha_n := \frac{\mathrm{d}K_{n+1}(X_0, \dots, X_n, \cdot)}{\mathrm{d}K_n(X_1, \dots, X_n, \cdot)}(X_{n+1}),$$

then for $\alpha_n^* := \sup_{\omega \in \Omega} |\alpha_n(\omega)|$ we assume

$$\lim_{n \to \infty} n^{2+\delta} \cdot \alpha_n^* = 0,$$

for any $\delta > 0$.

Remark 2.2. Our condition is similar to Condition 4.1 of Orey (1984). Orey assumes

$$\sum_{n=1}^{\infty} \alpha_n^* < \infty$$

and shows that under additional regularity assumptions for such measures P there holds a Shannon–McMillan-type theorem and that the sequence $(P \circ L_n^{-1})_{n \in \mathbb{N}}$ satisfies a large deviation principle with the specific relative entropy as rate function.

Note that our condition does not imply the stationarity of P.

In the Markovian case $P = \mu \otimes K^{\otimes \infty}$ the condition is satisfied because then all α_n are equal to zero.

If P is a m-step-Markov chain, i.e. $P = \mu \otimes K_1 \otimes K_2 \otimes \cdots \otimes K_m^{\otimes \infty}$, then $\alpha_n^* = 0$ for all $n \geq m$.

Condition 2.1 implies that P is locally absolutely continuous w.r.t. the product measure $\tilde{P} := \pi^{\otimes \infty}$ and that for all $n \in \mathbb{N}$ we get

$$\begin{aligned} k_{n+1}(X_0, \dots, X_{n+1}) &= \frac{\mathrm{d}K_{n+1}(X_0, \dots, X_n, \cdot)}{\mathrm{d}K_n(X_1, \dots, X_n, \cdot)}(X_{n+1}) \frac{\mathrm{d}K_n(X_1, \dots, X_n, \cdot)}{\mathrm{d}\pi}(X_{n+1}) \\ &= (1 + \alpha_n) k_n(X_1, \dots, X_{n+1}). \end{aligned}$$

Thus we have for $n \geq 2$ (writing $l := \lfloor n/2 \rfloor := \sup\{k \in \mathbb{N} \colon k \leq n/2\}$)

$$\begin{aligned} \left.\frac{\mathrm{d}P}{\mathrm{d}\tilde{P}}\right|_{\mathscr{F}_n} &= m(X_0) \prod_{i=1}^{n} k_i(X_0, \dots, X_i) \\ &= m(X_0) \prod_{i=1}^{l} k_i(X_0, \dots, X_i) \prod_{i=l+1}^{n} k_l(X_{i-l}, \dots, X_i) \left(\prod_{i=l+1}^{n} \prod_{j=l+1}^{i} (1 + \alpha_{j-1}) \right) \end{aligned}$$

and with

$$\psi_n := \frac{m(X_0) \prod_{i=1}^{l-1} k_i(X_0, \dots, X_i)}{\prod_{i=1}^{l} k_l(X_{n-l+i}, \dots, X_n, X_0, \dots, X_{i-1})},$$

$$f_n := \prod_{i=l}^{n} k_l(X_{i-l}, \dots, X_i) \prod_{i=1}^{l} k_l(X_{n-l+i}, \dots, X_n, X_0, \dots, X_{i-1})$$

and

$$\gamma_n := \prod_{i=l+1}^{n} \prod_{j=l+1}^{i} (1 + \alpha_{j-1})$$

we obtain

$$\left.\frac{\mathrm{d}P}{\mathrm{d}\tilde{P}}\right|_{\mathscr{F}_n} = \psi_n \cdot f_n \cdot \gamma_n,$$

i.e. the density allows a factorization, where f_n is shift-invariant for all n and γ_n converges to 1 as n goes to infinity. This will be of major importance in the proof of our main result.

The not-shift-invariant and "not-vanishing" part of the density, ψ_n, is needed to define $\Phi_n : \mathscr{M}_1(\Omega) \to \mathscr{M}_1(\Omega)$ for given $Q \in \mathscr{M}_1(\Omega)$, $A \in \mathscr{F}$ by

$$\Phi_n(Q)[A] := \frac{\int \psi_n \circ T_n^l 1_A \circ T^l \mathrm{d}Q}{\int \psi_n \circ T_n^l \mathrm{d}Q}$$

We will show below (Corollary 3.8) that the limit $\Phi : \mathscr{M}_1^S(\Omega) \to \mathscr{M}_1(\Omega)$, defined by

$$\Phi(Q)[A] := \lim_{n\to\infty} \Phi_n(Q)[A] \quad \text{for } Q \in \mathscr{M}_1^S(\Omega),\ A \in \mathscr{F},$$

exists and that it is $\mathscr{B}(\mathscr{M}_1^S(\Omega)) - \mathscr{B}(\mathscr{M}_1(\Omega))$-measurable. The crucial point in the proof of our theorem will be that $\Phi(L_n)[A]$ is a good approximation for the conditional probability of $A \in \mathscr{F}$, given the empirical process L_n.

We recall that the sequence $(P \circ L_n^{-1})_{n\in\mathbb{N}}$ satisfies a *large deviation principle* with rate function I, $I : \mathscr{M}_1(\Omega) \to [0, \infty]$, if

(i) I is lower semicontinuous.

(ii) I has compact level sets, i.e. for all $r \in \mathbb{R}$ the sets K_r,

$$K_r := \{Q \in \mathscr{M}_1(\Omega) : I(Q) \le r\},$$

are compact in $\mathscr{M}_1(\Omega)$.

(iii) For all $A \in \mathscr{B}(\mathscr{M}_1(\Omega))$ we have

$$\limsup_{n\to\infty} \frac{1}{n} \log P[L_n \in A] \le -\inf_{Q \in \mathrm{cl}\, A} I(Q),$$

$$\liminf_{n\to\infty} \frac{1}{n} \log P[L_n \in A] \ge -\inf_{Q \in \mathrm{int}\, A} I(Q),$$

where $\mathrm{cl}\, A$ and $\mathrm{int}\, A$ denote the closure resp. the interior of A.

We define upper- resp. lower-semicontinuous versions of the functional F as

$$\bar{F}(Q) := \inf_{U \text{ open},\, Q \in U} \sup_{\tilde{Q} \in U} F(\tilde{Q}),$$

$$\underline{F}(Q) := \sup_{U \text{ open},\, Q \in U} \sup_{\tilde{Q} \in U} F(\tilde{Q}).$$

The main result is the following theorem.

Theorem 2.3. *If*

1. $P \in \mathscr{M}_1(\Omega)$ *satisfies Condition* 2.1,
2. *the sequence* $(P \circ L_n^{-1})_{n \in \mathbb{N}}$ *satisfies a large deviation principle with rate function* I,
3. *for all sufficiently large* $n \in \mathbb{N}$: $P[F(L_n) = -\infty] < 1$, *and*
4. $\sup_{Q \in \mathscr{M}_1(\Omega)} (\bar{F}(Q) - I(Q)) = \sup_{Q \in \mathscr{M}_1(\Omega)} (\underline{F}(Q) - I(Q)) > -\infty$,

Then

$$K := \{Q \in \mathscr{M}_1(\Omega): \bar{F}(Q) - I(Q) = \sup_{\tilde{Q} \in \mathscr{M}_1(\Omega)} (\bar{F}(\tilde{Q}) - I(\tilde{Q}))\}$$

is nonempty and compact.

The family $(\hat{P}_n)_{n \in N}$ *is tight in* $\mathscr{M}_1(\Omega)$ *and any accumulation point* $\hat{P}$ *of the sequence has a representation*

$$\hat{P} = \int_K \Phi(Q) R(\mathrm{d}Q),$$

where $R \in \mathscr{M}_1(\mathscr{M}_1(\Omega))$ *and* $R(K) = 1$.

Remark 2.4. If the assumptions of the theorem are satisfied and K contains just a single point, i.e. $K = \{Q\}$, the theorem implies the convergence of the sequence $(\hat{P}_n)_{n \in \mathbb{N}}$ to $\Phi(Q)$.

Remark 2.5. Assumption 4 of Theorem 2.3 means for the case

$$F(Q) := -\infty \cdot 1_{\Gamma^c}(Q),$$

that $\Gamma \subseteq \mathscr{M}_1(\Omega)$ satisfies the following condition:

$$\inf_{Q \in \operatorname{int} \Gamma} I(Q) = \inf_{Q \in c\Gamma} I(Q),$$

i.e. one has the identity

$$\lim_{n \to \infty} \frac{1}{n} \log P[L_n \in \Gamma] = -\inf_{Q \in \Gamma} I(Q),$$

which corresponds to the so-called Sanov-property in Csiszar (1984).

Remark 2.6. If the functional F is real-valued and continuous, then conditions 3. and 4. of Theorem 2.3 are satisfied; examples are

$$F(Q) := \int f \,\mathrm{d}Q,$$

for $f: \Omega \to \mathbb{R}$ bounded and continuous, or

$$F(Q) = \int V(\omega, \eta) Q(\mathrm{d}\omega) Q(\mathrm{d}\eta),$$

where $V: \Omega^2 \to \mathbb{R}$ is bounded and continuous.

Remark 2.7. If one has a large deviation result in the finer τ-topology, i.e. the topology which is generated by the integrals over bounded, measurable functions, one can relax Condition 2.1: then the measurability of the densities k_n is sufficient; one does not have to impose their continuity, cf. the remark in the following

Example 2.8. Application to Markov Chains.

Our theorem easily applies to the following Markovian situation, for which we derive a level-III-version of Theorem 2.4 of Bolthausen and Schmock (1989). Moreover, we can give a more explicit representation for the accumulation points than in our quite general theorem. Let $\mu \in \mathscr{M}_1(S)$ be the initial distribution of the chain, K denotes a Markov transition kernel on $(S, \mathscr{B}(S))$. On μ and K we put the following

Condition. There exists a reference measure $\pi \in \mathscr{M}_1(S)$, such that

(i) π is invariant w.r.t. K, i.e. $\pi K = \pi$,
(ii) μ and K have continuous densities w.r.t. π, i.e.

$$\mu(\mathrm{d}x) = m(x)\pi(\mathrm{d}x), \qquad K(x, \mathrm{d}y) = k(x, y)\pi(\mathrm{d}y),$$

and m and k are continuous.
(iii) Let k be bounded and bounded away from zero.
(iv) Let m be bounded.

Remark. The main difference to Condition 2.2 of Bolthausen and Schmock (1989) is the assumed continuity of the densities k and m. We need this condition because we work in the weak topology, since we use large deviation results in this topology. Therefore we need the continuity of Φ in the weak topology for the proof of Proposition 3.9. If one has a large deviation result in the finer τ-topology, then, as in Bolthausen and Schmock (1989), one only needs the continuity of Φ w.r.t. this finer topology and therefore only the measurability of the densities k and m.

For the Markov chain $P = \mu \otimes K^{\otimes \infty}$, the specific relative entropy $h(Q \mid P)$ of a measure $Q \in \mathscr{M}_1(\Omega)$ w.r.t. P is well-defined by

$$h(Q \mid P) := \begin{cases} \lim\limits_{n \to \infty} \dfrac{1}{n} H_n(Q \mid P) & \text{if } Q \in \mathscr{M}_1^S(\Omega), \\ \infty & \text{otherwise,} \end{cases}$$

where $H_n(Q \mid P)$ denotes relative entropy w.r.t. $\mathscr{F}_n$, i.e.

$$H_n(Q \mid P) := \begin{cases} \int \log \dfrac{\mathrm{d}\pi_{n+1} Q}{\mathrm{d}\pi_{n+1} P} \mathrm{d}\pi_{n+1} Q & \text{if } \pi_{n+1} Q \ll \pi_{n+1} P, \\ \infty & \text{otherwise.} \end{cases}$$

The Markov chain satisfies Condition 2.1. Exercises 2.1.21 and 3.2.22 of Deuschel and Stroock (1989) together with Theorem 3.2 of Bolthausen and Schmock (1989), i.e., a projective limit argument imply that the sequence $(P \circ L_n^{-1})_{n \in \mathbb{N}}$ satisfies a large deviation principle with rate function $h(\cdot \mid P)$. Hence Theorem 2.3 applies with the specific relative entropy as rate function. Moreover we can compute the limit Φ in this special case:

For $l = \lfloor n/2 \rfloor$, $Q \in \mathscr{M}_1^S(\Omega)$, $A \in \mathscr{F}$, Φ is defined by

$$\Phi(Q)(A) = \lim_{n \to \infty} \Phi_n(Q)(A) = \lim_{n \to \infty} \frac{\int \psi_n \circ T_n^l 1_A \circ T^l \mathrm{d}Q}{\int \psi_n \circ T_n^l \mathrm{d}Q}.$$

But here

$$\psi_n \circ T_n^l = \frac{m(X_l) \prod_{i=1}^{l-1} k(X_{i-1}, X_i)}{\prod_{i=1}^{l} k(X_{i-1}, X_i)} = \frac{m(X_l)}{k(X_{l-1}, X_l)}.$$

Using the stationarity of Q we get for all $n \geq 2$

$$\begin{aligned} \Phi_n(Q)(A) &= \int \frac{m(X_l)}{k(X_{l-1}, X_l)} 1_A \circ T^l \mathrm{d}Q \Big/ \int \frac{m(X_l)}{k(X_{l-1}, X_l)} \mathrm{d}Q \\ &= \int \frac{m(X_1)}{k(X_0, X_1)} 1_A \circ T \mathrm{d}Q \Big/ \int \frac{m(X_1)}{k(X_0, X_1)} \mathrm{d}Q. \end{aligned}$$

Hence, any accumulation point $\hat{P}$ of $\hat{P}_n$ has a representation

$$\hat{P}[A] = \int_K \left(\int \frac{m(X_1)}{k(X_0, X_1)} 1_A \circ T \mathrm{d}Q \Big/ \int \frac{m(X_1)}{k(X_0, X_1)} \mathrm{d}Q \right) R(\mathrm{d}Q),$$

$A \in \mathscr{F}$.

Remarks. Since we consider transformations on process level, the accumulation points are, in contrast to the result of Bolthausen and Schmock (1989), no longer mixtures of Markov chains.

If P is a m-step-Markov chain, $P = \mu \otimes K_1 \otimes K_2 \otimes \cdots \otimes K_m^{\otimes \infty}$, one can compute the functional Φ in the same way as above. Then $\Phi_n = \Phi$, for all $n \geq 2m$.

If P is a stationary product measure, $P = \mu^{\otimes \infty}$, then $\Phi(Q) = Q$, i.e. in this case the accumulation points of $(\hat{P}_n)_{n \in \mathbb{N}}$ are mixtures of measures of K.

3. Proof of Theorem 2.3

The proof of Theorem 2.3 consists of two parts, illustrating the role of the two main ingredients for our result.

The first part states consequences of the large deviation principle and the regularity assumptions on F and makes essential use of Laplace's method (Varadhan's theorem). The proofs of this part do not depend on the structure of the measure P which is given in Condition 2.1.

In the second part we make essential use of functionals describing conditional probabilities, given the empirical process. Here it is Condition 2.1, which makes things work. We define functionals, depending on n, which describe the conditional expectation of the process P, given the empirical process L_n: Thanks to our fading condition there exists a sequence of l-step Markov chains which give a good approximation for P and there are functionals, describing the conditional evolution of these l-step Markov chains, given L_n, which on the other hand approximate the above functionals. The fading condition moreover implies the convergence of the functionals belonging to the Markov chains and with the help of their limit one can give a representation for the accumulation points of the sequence $\hat{P}_n$.

3.1. Consequences of the large deviation principle

The proofs in this section will be omitted for they carry over from Bolthausen and Schmock (1989) with obvious changes

Lemma 3.1. *The set* K,

$$K := \{\hat{Q} \in \mathscr{M}_1(\Omega) \colon \bar{F}(\hat{Q}) - I(\hat{Q}) = \sup_{Q \in \mathscr{M}_1(\Omega)} (\bar{F}(Q) - I(Q))\},$$

is nonempty and compact.

Proof. See Lemma 3.5 in Bolthausen and Schmock (1989). □

A version of Varadhan's theorem is as follows.

Lemma 3.2. *Let* $G \colon \mathscr{M}_1(\Omega) \to [-\infty, \infty)$ *be* $\mathscr{B}(\mathscr{M}_1(\Omega))$*-measurable and bounded above. If* $\bar{G}$ *and* $\underline{G}$ *denote the upper-, resp. lower- semicontinuous versions of* G, *then*

$$\limsup_{n \to \infty} \frac{1}{n} \log E[\exp(n \cdot G(L_n))] \leq \sup_{Q \in \mathscr{M}_1(\Omega)} (\bar{G}(Q) - I(Q))$$

and

$$\liminf_{n \to \infty} \frac{1}{n} \log E[\exp(n \cdot G(L_n))] \geq \sup_{Q \in \mathscr{M}_1(\Omega)} (\underline{G}(Q) - I(Q)).$$

Proof. See Corollary 3.3 in Bolthausen and Schmock (1989). □

As a consequence of Lemma 3.2, the next lemma shows that the distribution of the empirical process under the transformations $(\hat{P}_n)_{n \in \mathbb{N}}$ finally concentrates on the compact set K.

Lemma 3.3. *Let* R_n *be the distribution of* L_n *under* $\hat{P}_n$, *i.e.* $R_n \in \mathscr{M}_1(\mathscr{M}_1(\Omega))$ *is defined by*

$$R_n := \hat{P}_n \circ L_n^{-1},$$

then for every open neighbourhood U of K:

$$\lim_{n\to\infty} R_n[U^C] = 0.$$

Proof. See Lemma 4.4 in Bolthausen and Schmock (1989). □

Corollary 3.4. $\{R_n\colon n \in \mathbb{N}\} \subseteq \mathcal{M}_1(\mathcal{M}_1(\Omega))$ *is tight.*

Proof. See Corollary 4.5 (a) in Bolthausen and Schmock (1989). □

3.2. *Tightness of the transformed process*

Given $k, n \in \mathbb{N}$ and $l = \lfloor n/2 \rfloor$ we define the functional $\tilde{\Phi}_n^k\colon \mathcal{M}_1(\Omega) \to \mathcal{M}_1(S^k)$ by

$$\tilde{\Phi}_n^k(Q)[A] := \frac{\int (\psi_n \cdot \gamma_n) \circ T_n^l 1_A(X_l \cdots X_{l+k-1}) \mathrm{d}Q}{\int (\psi_n \cdot \gamma_n) \circ T_n^l \mathrm{d}Q},$$

for $Q \in \mathcal{M}_1(\Omega)$, $A \in \mathcal{B}(S^k)$, and now we can state a central proposition, concerning the conditional expectation of the process P, given the empirical process L_n, which is an analogon to Lemma 4.2 of Bolthausen and Schmock (1989) in our more general setting.

Proposition 3.5. *If* $k \in \mathbb{N}$, $A \in \mathcal{B}(S^k)$, *then for* $n \geq 2k$

$$E[1_A(X_0 \ldots X_{k-1}) \mid L_n] = \tilde{\Phi}_n^k(L_n)[A].$$

Proof. $\tilde{\Phi}_n^k(\cdot)[A]$ is $\mathcal{B}(\mathcal{M}_1(\Omega))$-measurable with Lemma 2.1 of Bolthausen and Schmock (1989), therefore the right-hand side is L_n-measurable. Let $C \in \sigma(L_n)$. We have to show that

$$E[1_A(X_0 \cdots X_{k-1}) 1_C] = E[\tilde{\Phi}_n^k(L_n)[A] 1_C].$$

We first define the (noncyclic) empirical process

$$\tilde{L}_n(\omega) := \frac{1}{n+1} \sum_{k=0}^{n} \delta_{T_n^k(\omega)}.$$

Note that $\tilde{L}_n$ is, in contrast to L_n, not $\mathcal{F}_n - \mathcal{B}(\mathcal{M}_1(\Omega))$-measurable. We now make use of the fact that $\tilde{P}$, 1_C, f_n, L_n and $\tilde{L}_n$ are cyclic shift-invariant, i.e.

$$\tilde{P} \circ T_n^{-1} = \tilde{P};\quad 1_C \circ T_n = 1_C;\quad f_n \circ T_n = f_n;\quad \tilde{L}_n \circ T_n = \tilde{L}_n;\quad L_n \circ T_n = L_n$$

and therefore

$$E[1_A(X_0 \cdots X_{k-1}) 1_C]$$

$$= \int \psi_n \gamma_n f_n 1_A(X_0 \cdots X_{k-1}) 1_C \mathrm{d}\tilde{P}$$

$$= \int \psi_n \gamma_n f_n 1_A(X_0 \cdots X_{k-1}) 1_C \frac{1}{n+1} \sum_{j=0}^{n} \mathrm{d}\tilde{P} \circ T_n^{-j}$$

$$= \int \frac{1}{n+1} \sum_{j=0}^{n} (\psi_n \gamma_n 1_A(X_0 \cdots X_{k-1})) \circ T_n^j f_n 1_C \mathrm{d}\tilde{P}$$

$$= \int \left(\int (\psi_n \gamma_n 1_A(X_0 \cdots X_{k-1}) \mathrm{d}\tilde{L}_n \right) f_n 1_C \mathrm{d}\tilde{P}$$

$$= \int \left(\int (\psi_n \gamma_n) \circ T_n^l 1_A(X_l \cdots X_{l+k-1}) \mathrm{d}\tilde{L}_n \right) f_n 1_C \mathrm{d}\tilde{P}.$$

With $(\psi_n \gamma_n) \circ T_n^l$ and, because of $n \geq l + k - 1$, also $1_A(X_l, \ldots, X_{l+k-1})$, being $\mathscr{F}_n$-measurable, one has

$$\int (\psi_n \gamma_n) \circ T_n^l 1_A(X_l \cdots X_{l+k-1}) \mathrm{d}\tilde{L}_n = \int (\psi_n \gamma_n) \circ T_n^l 1_A(X_l \cdots X_{l+k-1}) \mathrm{d}L_n$$

and with the definition of $\tilde{\Phi}_n^k$ we get

$$\int \left(\int (\psi_n \gamma_n) \circ T_n^l 1_A(X_l \cdots X_{l+k-1}) \mathrm{d}L_n \right) f_n 1_C \mathrm{d}\tilde{P}$$

$$= \int \left(\int (\psi_n \gamma_n) \circ T_n^l \mathrm{d}L_n \right) \tilde{\Phi}_n^k(L_n)[A] f_n 1_C \mathrm{d}\tilde{P}$$

$$= \int \left(\int (\psi_n \gamma_n) \circ T_n^l \mathrm{d}\tilde{L}_n \right) \tilde{\Phi}_n^k(L_n)[A] f_n 1_C \mathrm{d}\tilde{P}$$

$$= \int \left(\frac{1}{n+1} \sum_{j=0}^{n} (\psi_n \gamma_n) \circ T_n^j \right) \tilde{\Phi}_n^k(L_n)[A] f_n 1_C \mathrm{d}\tilde{P}$$

$$= \int \psi_n \gamma_n \tilde{\Phi}_n^k(L_n)[A] f_n 1_C \frac{1}{n+1} \sum_{j=0}^{n} \mathrm{d}\tilde{P} \circ T_n^{-j}$$

$$= \int \psi_n \gamma_n \tilde{\Phi}_n^k(L_n)[A] f_n 1_C \mathrm{d}\tilde{P}$$

$$= E[\tilde{\Phi}_n^k(L_n)[A] 1_C]. \qquad \square$$

For $k \in \mathbb{N}$ and $n \geq 2k$ $\Phi_n^k := \pi_k \circ \Phi_n : \mathscr{M}_1(\Omega) \to \mathscr{M}_1(S^k)$ is given by

$$\Phi_n^k(Q)[A] = \frac{\int \psi_n \circ T_n^l 1_A(X_l, \ldots, X_{l+k-1}) \mathrm{d}Q}{\int \psi_n \circ T_n^l \mathrm{d}Q},$$

where $Q \in \mathscr{M}_1(\Omega)$, $A \in \mathscr{B}(S^k)$, $l = \lfloor n/2 \rfloor$.

Note that Φ_n^k describes the conditional expectation of a l-step-Markov chain, given the empirical process, i.e. for $P = \mu \otimes K_1 \otimes \cdots \otimes K_{l-1} \otimes K_l^{\otimes \infty}$, one can state Proposition 3.5 with Φ_n^k instead of $\tilde{\Phi}_n^k$. The next lemma shows that one can approximate $\tilde{\Phi}$ by Φ.

Lemma 3.6. *For any $\varepsilon > 0$ and any $k \in \mathbb{N}$ there exists a $n_0 = n_0(k,\varepsilon)$, such that for all $n \geq n_0$, $A \in \mathscr{B}(S^k)$ and $Q \in \mathscr{M}_1(\Omega)$*

$$|\Phi_n^k(Q)[A] - \tilde{\Phi}_n^k(Q)[A]| < \varepsilon.$$

Proof. Let $\varepsilon < 1$ and k be given. With $l := \lfloor n/2 \rfloor$ and $\alpha^* := \max_{l \leq i \leq n-1} \alpha_i^*$ we get for $n \geq 2k$:

$$\begin{aligned}\gamma_n &= \prod_{j=l+1}^{n} \prod_{i=l+1}^{j} (1 + \alpha_{i-1}) \\ &\leq \prod_{j=l+1}^{n} \prod_{i=l+1}^{j} (1 + \alpha_{i-1}^*) \\ &\leq (1 + \alpha^*)^{l^2}.\end{aligned}$$

Now we choose n_0 according to Condition 2.1 such that $(1 + \alpha^*)^{l^2} \leq 1 + \varepsilon/4$ and $(1 - \alpha^*)^{l^2} \geq 1 - \varepsilon/4$ for $n \geq n_0$. Hence

$$1 - \frac{\varepsilon}{4} \leq \gamma_n \leq 1 + \frac{\varepsilon}{4}$$

for all $n \geq n_0$, which implies

$$\Phi_n^k(Q)[A] \cdot \frac{1 - \varepsilon/4}{1 + \varepsilon/4} \leq \tilde{\Phi}_n^k(Q)[A] \leq \Phi_n^k(Q)[A] \cdot \frac{1 + \varepsilon/4}{1 - \varepsilon/4}$$

for all $A \in \mathscr{B}(S^k)$ and all $Q \in \mathscr{M}_1(\Omega)$ and therefore

$$\Phi_n^k(Q)[A] - \frac{\varepsilon}{2} \leq \tilde{\Phi}_n^k(Q)[A] \leq \Phi_n^k(Q)[A] + \varepsilon. \quad \square$$

The next lemma shows the convergence of the sequence $(\Phi_n(Q))$ for stationary measures Q.

Lemma 3.7. *Given $\varepsilon > 0$ there exists a $n_0 \in \mathbb{N}$, such that for all $j \in \mathbb{N}$, $A \in \mathscr{F} = \mathscr{B}(\Omega)$ and for all $Q \in \mathscr{M}_1^S(\Omega)$*

$$|\Phi_{n_0}(Q)[A] - \Phi_{n_{0+j}}(Q)[A]| < \varepsilon.$$

Proof. Remember that Condition 2.1 says that $n^{2+\delta} \cdot \alpha_n^*$ is small as n tends to infinity. Hence, given $\varepsilon > 0$, we can choose n_1, such that $(1 + \alpha_l^*)^{l+1} \leq 1 + \varepsilon_l$ and $(1 - \alpha_l^*)^{l+1} \geq 1 - \varepsilon_l$ for all $n \geq n_1$, where $l := \lfloor n/2 \rfloor$ and

$$\varepsilon_m := \frac{\varepsilon}{4m^{1+\delta}}.$$

Let n_2 be large enough, that $\sum_{n=l}^{\infty} 1/n^{1+\delta} < 1$, where $l := \lfloor n_2/2 \rfloor$. Now we define $n_0 := \max\{n_1, n_2\}$ and set $l := \lfloor n_0/2 \rfloor$. Let $A \in \mathscr{F}$ and $Q \in \mathscr{M}_1^S(\Omega)$ be given.

We first give an estimate for the difference $\Phi_{n_0+1}(Q)[A] - \Phi_{n_0}(Q)[A]$:
If $\lfloor (n_0 + 1)/2 \rfloor = l$, then with the definition of ψ_n we have $\psi_{n_0+1} \circ T_{n_0+1}^l = \psi_{n_0} \circ T_{n_0}^l$ therefore $\Phi_{n_0+1}(Q)[A] = \Phi_{n_0}(Q)[A]$.

If $\lfloor (n_0+1)/2 \rfloor = l+1$, we get

$$\psi_{n_0+1} \circ T_{n_0+1}^{l+1} = \frac{m(X_{l+1}) \prod_{i=1}^{l} k_i(X_{l+1}, \ldots, X_{l+1+i})}{\prod_{i=0}^{l} k_{l+1}(X_i, \ldots, X_{i+l+1})}$$

$$= \frac{m(X_{l+1}) \prod_{i=1}^{l-1} k_i(X_{l+1}, \ldots, X_{l+1+i})}{\prod_{i=0}^{l-1} k_l(X_{i+1}, \ldots, X_{i+l+1})} \cdot \frac{1}{\prod_{i=0}^{l}(1+\alpha_l \circ T^i)}$$

$$= \psi_{n_0} \circ T_{n_0}^{l} \circ T \cdot \frac{1}{\prod_{i=0}^{l}(1+\alpha_l \circ T^i)}.$$

Hence

$$(\psi_{n_0} \circ T_{n_0}^{l} \circ T) \cdot \frac{1}{(1+\alpha_l^*)^{l+1}} \le \psi_{n_0+1} \circ T_{n_0+1}^{l+1} \le (\psi_{n_0} \circ T_{n_0}^{l} \circ T) \cdot \frac{1}{(1-\alpha_l^*)^{l+1}}.$$

Now we get

$$(1-\varepsilon_l) \cdot (\psi_{n_0} \circ T_{n_0}^{l} \circ T) \le \psi_{n_0+1} \circ T_{n_0+1}^{l+1} \le (1+2\varepsilon_l) \cdot (\psi_{n_0} \circ T_{n_0}^{l} \circ T).$$

The stationarity of Q gives

$$\Phi_{n_0}(Q)[A] = \frac{\int (\psi_{n_0} \circ T_{n_0}^{l} \circ T) \cdot (1_A \circ T^{n_0+1}) \mathrm{d}Q}{\int \psi_{n_0} \circ T_{n_0}^{l} \circ T \, \mathrm{d}Q}$$

and this implies

$$\Phi_{n_0}(Q)[A] \cdot \frac{1-\varepsilon_l}{1+2\varepsilon_l} \le \Phi_{n_0+1}(Q)[A] \le \Phi_{n_0}(Q)[A] \cdot \frac{1+2\varepsilon_l}{1-\varepsilon_l}.$$

Hence

$$\Phi_{n_0}(Q)[A] - 3\varepsilon_l \le \Phi_{n_0+1}(Q)[A] \le \Phi_{n_0}(Q)[A] + 4\varepsilon_l.$$

Induction and the triangle inequality give, for all $j \in \mathbb{N}$,

$$\Phi_{n_0}(Q)[A] - 3 \cdot \sum_{m=l}^{l+j-1} \varepsilon_m \le \Phi_{n_0+j}(Q)[A] \le \Phi_{n_0}(Q)[A] + 4 \cdot \sum_{m=l}^{l+j-1} \varepsilon_m.$$

Hence

$$\Phi_{n_0}(Q)[A] - \frac{3}{4}\varepsilon \le \Phi_{n_0+j}(Q)[A] \le \Phi_{n_0}(Q)[A] + \varepsilon. \qquad \square$$

Corollary 3.8.

(1) *If $n \ge 2$, then Φ_n is continuous on $\mathscr{M}_1(\Omega)$ and therefore $\mathscr{B}(\mathscr{M}_1(\Omega)) - \mathscr{B}(\mathscr{M}_1(\Omega))$-measurable.*

(2) *The limit $\Phi: \mathscr{M}_1^S(\Omega) \to \mathscr{M}_1(\Omega)$, defined by $\Phi(Q)[A] := \lim_{n\to\infty} \Phi_n(Q)[A]$, for $Q \in \mathscr{M}_1^S(\Omega)$, $A \in \mathscr{F}$, exists.*
Φ is continuous on $\mathscr{M}_1^S(\Omega)$ and therefore $\mathscr{B}(\mathscr{M}_1^S(\Omega)) - \mathscr{B}(\mathscr{M}_1(\Omega))$-measurable.

Proof. (1): If $n \geq 2$, then for $f \in C_b(\Omega)$ $Q \in \mathscr{M}_1(\Omega)$ and $l = \lfloor n/2 \rfloor$ we have

$$\int f \, \mathrm{d}\Phi_n(Q) = \frac{\int \psi_n \circ T_n^l f \circ T^l \mathrm{d}Q}{\int \psi_n \circ T_n^l \mathrm{d}Q}.$$

Since $\psi_n \in C_b(\Omega)$ Φ_n is continuous.

(2): Lemma 3.7 shows that the limit Φ exists. Its continuity follows from (1) and Lemma 3.7. □

The proof of the following proposition contains very much of what one may call the central idea for the "whole story", namely how to make use of functionals, describing conditional expectations, given the cyclic empirical process L_n.

Proposition 3.9. *If $\hat{P}_n^k := \pi_k(\hat{P}_n)$, then for all $k \in \mathbb{N}$ $(\hat{P}_n^k)_{n \in \mathbb{N}}$ is tight in $\mathscr{M}_1(S^k)$ and any weak limit point $\hat{P}^k$ has a representation*

$$\hat{P}^k = \int_{\mathscr{M}_1^S(\Omega)} \Phi^k(Q) \, R(\mathrm{d}Q),$$

where $R \in \mathscr{M}_1(\mathscr{M}_1(\Omega))$.

Proof. If $k \in \mathbb{N}$ then for $A \in \mathscr{B}(S^k)$, $n \geq 2k$ and $l = \lfloor n/2 \rfloor$ we obtain using Proposition 3.5

$$\begin{aligned}
\hat{P}_n^k[A] &= E[\exp(n \cdot F(L_n)) 1_A(X_0, \ldots, X_{k-1})]/E[\exp(n \cdot F(L_n))] \\
&= E[\exp(n \cdot F(L_n)) E[1_A(X_0, \ldots, X_{k-1}) | L_n]]/E[\exp(n \cdot F(L_n))] \\
&= E[\exp(n \cdot F(L_n)) \tilde{\Phi}_n^k(L_n)[A]]/E[\exp(n \cdot F(L_n))] \\
&= \int_\Omega \tilde{\Phi}_n^k(L_n)[A] \, \mathrm{d}\hat{P}_n \\
&= \int_{\mathscr{M}_1(\Omega)} \tilde{\Phi}_n^k(Q)[A] \, \hat{P}_n \circ L_n^{-1}(\mathrm{d}Q) \\
&= \int_{\mathscr{M}_1^S(\Omega)} \tilde{\Phi}_n^k(Q)[A] \, R_n(\mathrm{d}Q),
\end{aligned}$$

where in the last equality we make use of the fact that $R_n = \hat{P}_n \circ L_n^{-1}$ only puts mass to stationary measures. We now show the relative compactness of the family $(\hat{P}_n^k)_{n \in \mathbb{N}}$, which by applying Prohorov's theorem proves their tightness. Let $(\hat{P}_{n_j}^k)_{j \in \mathbb{N}}$ be a subsequence of $(\hat{P}_n^k)_{n \in \mathbb{N}}$. Using Corollary 3.4 we see that $(R_{n_j})_{j \in \mathbb{N}}$ has a converging subsequence, which we denote by $(R_{n_{j_i}})_{i \in \mathbb{N}}$, and its limit with R. Let $\varepsilon > 0$ and $f \in C_b(S^k)$ be given. Lemma 3.6 shows that there exists an $i_0 \in \mathbb{N}$, such that for all $i \geq i_0$

$$\left| \int_{\mathscr{M}_1(\Omega)} \int_{S^k} f \, \mathrm{d}\tilde{\Phi}_{n_{j_i}}^k(Q) \, R_{n_{j_i}}(\mathrm{d}Q) - \int_{\mathscr{M}_1(\Omega)} \int_{S^k} f \, \mathrm{d}\Phi_{n_{j_i}}^k(Q) \, R_{n_{j_i}}(\mathrm{d}Q) \right| < \varepsilon/3.$$

With Lemma 3.7 and Corollary 3.8 being valid in particular for the marginal distributions of Φ_n, we see that there is an $i_1 \in \mathbb{N}$, such that for all $i \geq i_1$

$$\left| \int_{\mathcal{M}_1^S(\Omega)} \int_{S^k} f \, d\Phi_{n_{ji}}^k(Q) \, R_{n_{ji}}(dQ) - \int_{\mathcal{M}_1^S(\Omega)} \int_{S^k} f \, d\Phi^k(Q) \, R_{n_{ji}}(dQ) \right| < \varepsilon/3.$$

As $\int_{S^k} f \, d\Phi^k \in C_b(\mathcal{M}_1^S(\Omega))$, there is an $i_2 \in \mathbb{N}$, such that for all $i \geq i_2$

$$\left| \int_{\mathcal{M}_1^S(\Omega)} \int_{S^k} f \, d\Phi^k(Q) \, R_{n_{ji}}(dQ) - \int_{\mathcal{M}_1^S(\Omega)} \int_{S^k} f \, d\Phi^k(Q) \, R(dQ) \right| < \varepsilon/3.$$

Hence for all $i \geq \max\{i_0, i_1, i_2\}$:

$$\begin{aligned} &\left| \int_{S^k} f \, d\hat{P}_{n_{ji}}^k - \int_{\mathcal{M}_1^S(\Omega)} \int_{S^k} f \, d\Phi^k(Q) \, R(dQ) \right| \\ &= \left| \int_{\mathcal{M}_1(\Omega)} \int_{S^k} f \, d\tilde{\Phi}_{n_{ji}}^k(Q) \, R_{n_{ji}}(dQ) - \int_{\mathcal{M}_1^S(\Omega)} \int_{S^k} f \, d\Phi^k(Q) \, R(dQ) \right| \\ &< \varepsilon, \end{aligned}$$

and this shows the convergence of $(\hat{P}_{n_{ji}}^k)_{i \in \mathbb{N}}$ to $\int_{\mathcal{M}_1^S(\Omega)} \Phi^k(Q) \, R(dQ)$. □

Corollary 3.10. $(\hat{P}_n)_{n \in \mathbb{N}}$ *is tight.*

Proof. This follows from the tightness of the marginal distributions. □

Now we can complete the proof of Theorem 2.3. If $(\hat{P}_{n_j})_{j \in \mathbb{N}}$ is any converging subsequence of $(\hat{P}_n)_{n \in \mathbb{N}}$ with limit measure $\hat{P}$, then the marginals $(\hat{P}_{n_j}^k)_{j \in \mathbb{N}} = (\pi_k(\hat{P}_{n_j}))_{j \in \mathbb{N}}$ also converge to $\hat{P}^k = \pi_k(\hat{P})$. By Corollary 3.4 we may assume that $(R_{n_j})_{j \in \mathbb{N}}$ converges to a limit R. According to Proposition 3.9, $\hat{P}^k$ has for all $k \in \mathbb{N}$ and all $A \in \mathcal{B}(S^k)$ a representation

$$\hat{P}^k[A] = \int_{\mathcal{M}_1^S(\Omega)} \Phi^k(Q)[A] \, R(dQ).$$

Hence for all $A \in \mathcal{F}$ we have

$$\hat{P}[A] = \int_{\mathcal{M}_1^S(\Omega)} \Phi(Q)[A] \, R(dQ).$$

By Lemma 3.3 we know that the mass of R is concentrated on the compact set K, therefore we finally obtain

$$\hat{P}[A] = \int_K \Phi(Q)[A] \, R(dQ),$$

which completes the proof of Theorem 2.3.

Acknowledgement

The authors would like to thank Prof. Hans Foellmer for proposing the problem to us and for many stimulating discussions and we would like to thank the referee for making a useful remark concerning our fading condition.

References

E. Bolthausen and U. Schmock, On the maximum entropy principle for uniformly ergodic Markov chains, Stochastic Processes Appl. 33 (1989) 1–27.

I. Csiszar, Sanov property, generalized I-projection and a conditional limit theorem, Ann. Probab. 12(3) (1984) 768–793.

I. Csiszar, T.M. Cover and B.S. Choi, Conditional limit theorems under Markov conditioning, IEEE Trans. Inform. Theory 33(6) (1987) 788–801.

J.D. Deuschel and D.W. Stroock, Large Deviations (Academic Press, New York, 1989).

H.O. Georgii, Large deviations and maximum entropy principle for interacting random fields on Z^d, Ann. Probab. 21(4) (1993) 1845–1875.

S. Orey, Large deviations in ergodic theory, in: Seminar on Stochastic Processes 1984 (Birkhäuser, Boston, 1986).

U. Schmock, On the maximum entropy principle for Markov chains and processes, Dissertation, TU (Berlin, 1990).